

# Review of: "Towards Responsible AI-Assisted Scholarship: Comparative Assessment of Generative Models and Adoption Recommendations"

Zaenal Arifin Siregar<sup>1</sup>

<sup>1</sup> National Research and Innovation Agency (BRIN)

Potential competing interests: No potential competing interests to declare.

Exhaustive examination of trends, particularly the intersection of AI and research, is indicative of a thorough comprehension of the topic and demonstrates quality research. Also commendable is the use of a mixed-method approach, which permits a multifaceted examination of the research problem.

However, it is necessary to resolve a few areas for improvement. First, the involvement of only two assessors in evaluating AI output raises concerns regarding the process's reliability and objectivity. Increasing the size of the evaluation council could enhance the validity of the study's findings.

Regarding Table 1 of the Key Findings, where Claude received the highest composite score (7.9), readers would benefit from a more thorough explanation of how this score was calculated and its significance.

In conclusion, while the statement "This demonstrates that competitive benchmarking can identify skill gaps based on training approaches" is insightful, it would be more persuasive if statistical results, such as ANOVA or Tukey tests, were presented. The statement that Claude scored substantially higher than LaMDA and Galactica ( $p0.05$ ), while there was no significant difference between Claude and Sydney, should be supported by statistical evidence.

Best regards.